DOCUMENT RESUME

ED 039 175 SP 003 841

AUTHOR Resnick, Lauren B; Kiss, Lois E.
TITLE Discrimination Training and Feedback in Shaping Teacher Behavior.
PUB DATE 70
NOTE 24p.; Paper presented at annual meeting, AERA, Minneapolis, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.30
DESCRIPTORS *Behavior Change, *Discrimination Learning, *Feedback, Microteaching, *Teacher Behavior, *Teacher Education, Teaching Models

ABSTRACT

In an experiment to test the effectiveness of discrimination training--contrasting good and poor teacher behaviors and demonstrating the stimulus occasions for these behaviors--as compared with feedback from an instructor in microteaching, a discrimination model for a limited class of teaching behaviors was devised and subjects were divided into three groups according to the type of training received. The first group received discrimination training, practice, and feedback (DPF); the second group received discrimination training and practice (DP); and the third group received practice and feedback (PF). A comparison of pre- and posttest scores on several behavior rating scales showed consistent and significant improvement of the DPF and DP groups as compared with slight improvement in the PF group. These results seem to indicate that discrimination training could replace feedback from an instructor in many cases and thus effect substantial economies in teacher education. (An appendix contains pretest posttest comparison graphs) (RT)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED039175

Discrimination Training and Feedback in Shaping Teacher Behavior[1]

by

Lauren B. Resnick & Lois E. Kiss

University of Pittsburgh

In learning a complex social skill such as teaching an individual must acquire both a new repertoire of behaviors and a set of discriminations as to when each specific behavior in that repertoire is appropriate. Studies in "microteaching" have shown that specified teaching behaviors can be effectively shaped or modified through a process of repeated practice interspersed with critical review by the trainee of his own videotaped performance (Allen and Ryan, 1969). A study by McDonald & Allen (1967, Chapter IV) showed that provision of "models" of effective performance of the teaching behaviors to be acquired enhances acquisition of those behaviors to some degree. However, the model's effectiveness was greatest when S viewed it along with E, who pointed out both the appropriate behaviors of the model and the occasions (student responses) suitable for those behaviors.

McDonald and Allen (1967, Chapter II) also found that Ss who had feedback sessions in which E helped them review videotapes of their performance showed significantly greater improvement than Ss who viewed their videotapes

---

[1] Paper prepared for presentation at 1970 meetings of the American Educational Research Association. The research reported here was supported by a grant by the Office of Education, U. S. Department of Health, Education and Welfare, to the Learning Research and Development Center, University of Pittsburgh. The points of view or opinions presented do not necessarily represent official Office of Education position or policy.

SP003841

alone. This effect of feedback from a supervisor remained even when pre-practice models were provided (McDonald and Allen, 1967, Chapter IV). In both the modeling and the feedback contexts, E's commentary apparently functioned as a form of informal "discrimination training," contrasting good and poor teacher behaviors and cueing the stimulus occasions for these behaviors.

The present study explores the effect of a special form of "discriminative modeling," in which a succession of model performances are deliberately chosen to contrast appropriate and inappropriate teacher behaviors and stimulus occasions. It is designed to test the hypothesis that when both responses and stimulus occasions have been adequately discriminated in advance of practice, Ss can apply the discriminations to their own subsequent behavior and thus learn to "self-edit" their own performance. This capability should make it possible for Ss to benefit from practice without need of direct feedback from a supervisor or experimenter.

Bandura, Ross and Ross (1963) have shown that when a variety of models of a class of behavior are shown, children can acquire the class of behavior without the idiosyncratic features of any single model's performance. Ss in effect learn to produce new exemplars of the behavior class through observing multiple models of the behavior. A later study (Bandura and Harris, 1966) showed that children could learn a new class of grammatical forms most effectively when models of the sentence form were interspersed with sentences of other types and the child's attention

directed to differences between "correct" and "incorrect" sentences. It is this use of contrasting, multiple models that we term "discriminative modeling," and that we seek to apply to adult's learning of a complex interactive repertorire.

Method

For purposes of the study, a limited class of teaching behaviors was examined. These are all behaviors required in effective testing or diagnosis of the intellectual skills of young children, where testing is conducted in a one-to-one "tutorial" type setting. The specific behaviors studied were chosen on the basis of detailed behavioral analyses of tutorial teaching skills (Resnick and Kiss, 1970).

Three elements of training were identified: 1) Discrimination Training (Variable D); 2) Practice (Variable P); and 3) Feedback (Variable F).

Discrimination training (Variable D) consisted of viewing, rating and discussing videotaped tutorial diagnostic sessions. Discrimination tapes were constructed of "model sessions" interspersed with "problem sessions," in order to contrast appropriate and inappropriate behaviors by a teacher. Problem sessions were chosen to illustrate behaviors which caused lengthy or inaccurate testing.

For Practice sessions (Variable P) Ss came individually or in pairs to a special laboratory. During practice sessions Ss (teacher trainees) diagnosed several children on each of several related tasks. Each interaction was recorded on videotape, and after completing her ratings for a child, S could view the tape before beginning work with another child.

Feedback (Variable F) refers to feedback from the experimenter. In the two groups receiving feedback E served as evaluator, rating the trainee's performance and discussing the video playback during the practice sessions. Practice sessions for these groups were scheduled individually. The group not receiving feedback practiced in pairs, alternating as tester and observer. They were free to view videotapes together and discuss each other's performance.[2] The experimenter, however, did not interact with them.[3]

Varying the three elements of training produced three experimental groups. One group (labelled DPF) received all three training elements, Discrimination, Practice and Feedback for the experimenter. A second group (PF) had Practice and Feedback, but no Discrimination training. The third group (DP) had Discrimination Training and Practice, but no Feedback.

(SLIDE 1) This slide summarizes the entire experimental procedure for each of the three groups. All groups began with an orientation session, in which they were familiarized with the purpose of the training program, the criteria that would be used in judging their performance, and the materials they would be using in working with the children. All groups were then pretested. For the pretest, each trainee diagnosed the performance of five or six children on two tasks. The first task was a "multiple dis-

---

[2]Limits on time made it necessary for Ss in this treatment to share the number of children available for testing during a given practice session. Thus, these Ss each tested only 3 children rather than the 6 tested by Ss in Feedback treatments.

[3]Informal observations suggest that trainees working in pairs actually engaged in very little discussion of each other's performance.

crimination," which called for the child to point out nine basic shapes when each was named. The second task was a "two-part discrimination" and called for the child to distinguish the numerals "6" and "9". Ss later received discrimination and practice in diagnosing the two classes of tasks, multiple discriminations and two-part discriminations, but did not practice the particular tasks used in the pre- and post-test sessions.

The children tested were 4-6 years old, drawn from Headstart and Kindergarten classes at a nearby elementary school. A remote-controlled camera recorded the pre- and post-test interactions for later scoring purposes, but the videotapes were not shown to trainees.

A group discrimination training session came next for Ss in groups DPF and DP. The training tape for this session focused on multiple discrimination tasks. Ss in all groups then had a practice session in which they tested children on multiple discrimination tasks. Two of the groups (DPF and PF) had experimenter feedback during the practice session. The cycle of discrimination training and practice was then repeated, this time focusing on the testing of two-part discriminations. Finally, all Ss were post-tested in an exact replication of the pre-test procedure.

Results

Pre- and post-tests were scored on two general measures of diagnostic competence and on a series of specific trainee behaviors.

(SLIDE 2) The first general measure was a Certainty rating, designed to assess the confidence with which independent observers could judge the child's ability to perform each of the tasks tested. Using the form shown

here and working from videotape playbacks, two independent observers tallied the child's responses to each item in the test and then rated their certainty that a child could or could not perform the objective. A rating of 3+ or 3- indicated a high certainty; 1- or 1+ or 0 a very low certainty. In averaging certainty ratings, signs were ignored. This resulted in a measure that gives equal weight to strong certainty that the child can perform the task and strong certainty that he cannot. For each trainee separate scores were computed for multiple discrimination tasks and for two-part discrimination tasks, averaging across the children she tested in a pre- or post-test session.

(SLIDE 3)These graphs show the changes in pre- and post-test means for the three experimental groups on the Certainty measure. The data for this and all subsequent measures were analyzed using multiple covariance analysis (Cooley & Lohnes, 1970), a procedure which compares post-test means which have been adjusted for pre-test level, and which also permits simultaneous examination of sets of related measures in order to detect common trends. This capacity permitted the examination of scores for multiple and two-part discriminations combined, as well as of scores for each type of task separately. For this measure, there were no significant differences between groups, although differences were in the expected directions.

(SLIDE 4) The same finding is repeated for the second general measure, the average number of minutes spent with a child on a single task. All groups showed substantial reductions in the time required to complete a

diagnosis; and all groups also showed a significant reduction in variability from pre- to post-test. However, there was no differential effect for the three groups. Thus, general diagnostic skill was apparently not differentially affected by the three training conditions used in this experiment.

The next five measures are all based on direct observation of specific component behaviors that were treated in the discrimination training tapes. Inter-observer reliability for these measures ranged from 80 to 100 percent.

(SLIDE 5) During the pretest, many trainees tended to eliminate some of the choice objects during the testing session, so that a child was choosing his response from a group of three or four rather than of nine or eleven objects. Often trainees eliminated some of the choices before retesting, making the retest ineffective because of the increased chance of a random response being correct. The slide shows changes from pre- to post-test in the behavior of eliminating choices. Differences between the groups were significant for the two tasks combined ($F = 9.76$; $df = 2,23$; $p < .01$) and for the multiple discrimination task alone ($F = 7.26$; $df = 2,23$; $p < .01$). As you can see, on the multiple discrimination task the behavior was reduced nearly to zero in the two groups that had discrimination training, while the group without discrimination training actually increased its rate of elimination behavior.

(SLIDE 6) This is a measure of the number of times the trainee permitted the child to eliminate choice stimuli by picking up or putting aside each object as he dealt with it. Differences among treatment groups were significant

for the two tasks combined ($F=3.44;df=2,23;p<.05$). The behavior occurred mainly in the multiple discrimination task. Although the large pre-test differences between the groups require that the post-test effects be interpreted with caution, the differences are in the direction predicted for behaviors dealt with in discrimination training. Occurences were reduced nearly to zero for the groups with discrimination training (DP and DPF). Occurrence of the behavior actually increased on the post-test for the group without discrimination training. This increase for the DP group illustrates the limitations of relying on feedback during practice sessions as a major training technique. If a trainee did not happen to test a child who eliminated materials in training, she did not know to handle the situation when it arose in the post-test. The group that considered the situation in discrimination training, however, knew how to deal with it even if it had not arisen in practice sessions.

(SLIDE 7) This is a measure of "cueing," defined as all physical indications a trainee made that might tend to influence a child's choice in responding to a request. Behaviors scored as cueing included adding, pointing to, tapping, moving forward, or otherwise indicating a specific stimulus to the child just before of after requesting a response. Differences between groups were significant for the two tasks combined ($F=7.51;df=2,23;p<.01$) and for the multiple discrimination task alone ($F=6.46;df=2,23;p<.01$). The two groups undergoing discrimination training reduced the amount of cueing in their post-test performances nearly to zero. The group which did not

experience discrimination training (PF) was cueing nearly eight times as often as the others on the multiple discrimination post-test, although it had begun with a slightly lower rate on the pretest.

(SLIDE 8) This is a measure of "prompting" behavior, defined as any verbal information the trainee gave the child which might direct him to choose the right response or the wrong response, or to change responses. A "prompt" was scored whenever a trainee named the stimulus objects for the child, or gave him information about the object she wanted him to choose (e.g. "Show me a triangle; it's like a roof"); or when the trainee asked for an object the child was holding or touching, asked for a piece no longer on the table, or asked for numbers in order; or when the trainee provided feedback on incorrect responses by saying "no" or "try again." Group differences are significant for the two tasks combined ($F=4.44$;$df=2,23$: $p < .05$). Comparison of the slopes in the graphs suggest that for each task the group without discrimination training (PF) improved somewhat less than the other two. However, the higher prompting rate of group DP on the pretest, especially in the multiple discrimination, requires that these results be interpreted with caution.

(SLIDE 9) Determining a child's ability to distinguish between two related stimuli requires some elements in the test which are used to distract the child from the two relevant responses and force him to make a new response each time he is asked rather than following a "back and forth" pattern. Two different measures of distracting behavior by the trainee are represented here. They were used only on two-part tasks. For the top

measure, Failing to Add Distracting Stimuli, the observer checked each time the trainee used only the two relevant stimuli (i.e., "6" and "9") during some part of the task. For the bottom measure, Failing to Distract Child between Responses, the observer checked each time the trainee re-tested both items without distracting the child between cycles either by asking for an irrelevant response or by shuffling the position of the materials. Thus, these are measures of a trainee's failure to distract appropriately.

Differences between the treatment groups are significant for the two measures considered simultaneously ($F=5.10; df=2,23; p<.05$). Comparison of slopes shows that the group without discrimination training (PF) improved less than the other two (DPF and DP). For the top measure, Failing to Add Distracting Stimuli, however, the PF group began with a near zero occurence and thus had little room for improvement. Therefore, the differences in slope are probably attributable to the wide pre-test differences, rather than to post-test differences. Pre-test differences are less marked for the bottom measure, Failing to Distract Child, and post-test differences are sharper. While the PF group exhibits the behavior on only 12 percent of tasks on the post-test, even such a relatively low occurence could seriously affect the accuracy of testing. The two groups that received discrimination training both completely eliminated the Failure to Distract behavior.

Each of the behaviors just discussed had been specifically treated in the Discrimination training tapes. In each case, the two groups that saw the tapes (DPF and DP) improved significantly more than the group that

did not see the tapes, while feedback from the experimenter offered no advantage. This finding supports an interpretation of discrimination pre-training as establishing conditions which make it possible for the trainee to "edit" his own behavior, thus decreasing reliance on judgments by others.

Since there was no comparable effect of discrimination training on the general measures of Certainty and Time, the effect appears to be specific to the behaviors focussed on in pre-training. For other behaviors, those for which the "self-editing" skill would not have been developed, it would be expected that expert feedback would be essential to improvement. One of the behaviors measured in the pre and post-tests allows us to test this inference, since there was no explicit attention to it in the discrimination training tapes.

(SLIDE 10) In this measure, occasions on which a retest should have been made but was not were counted. Occasions requiring retesting are those in which the response is prompted or cued or the response might be correct by chance. The discrimination training tape did not focus on the retesting skill explicitly although it was touched upon tangentially. The measure was derived from the observer's certainty rating forms. A count was made of the items rated less than "3" by one or more observers where no retest was performed by the trainee.

Differences between the experimental groups were significant for the two measures combined ($F=4.84; df=2,23; p < .05$) and for the two-part discrimination considered separately ($F=3.77; df=2,23; p < .05$). In this case, the group without feedback showed no improvement on either type of task, while the other two groups showed equivalent reductions in means (taking pretest

performance into account). Thus, where pre-training has not explicitly developed the self-editing capacity, feedback appears to be the critical variable in improving performance.

DISCUSSION

The results of this experiment support the view that discriminative modeling techniques can produce a capacity for "self-editing" which substantially reduces a trainee's reliance on outside feedback. The study demonstrates that modeling techniques can be effectively used in developing interactive skills in adults. In practical terms, discrimination pretraining offers the possibility of doing away with the costly one-to-one feedback that has been found to be a critical requirement in other microteaching experiments. Using multiple models on film or videotape, discrimination sessions could be provided for groups at a central location; trainees could then practice in times and places convenient to them.

Although promising, the technique of discriminative pretraining requires further investigation. Most important, it is necessary to determine whether effects on specific component behaviors can be cumulated to produce a difference in general performance. Further experiments are planned to test directly the hypothesis that if enough specific components

are acquired and organized into a smooth performance, general measures will also register an improvement. If such an improvement can be effected through the use of discriminative modeling without direct feedback, substantial economy and improved effectiveness in teacher training efforts can be anticipated.

References

Allen, D. W. & Ryan, K. Microteaching. Reading, Massachusetts: Addison-Wesley, 1969.

Bandura, A., Ross, D., & Ross, S. A. Imitation of film-mediated aggressive models. Journal of Abnormal Social Psychology, 1963, 66, 3-11.

Bandura, A. & Harris, M. B. Modification of syntactic style. Journal of Experimental Child Psychology, 1966, 4, 341-352.

Cooley, W. W. & Lohnes, P. R. Multivariate Data Analysis. New York: John Wiley, in press.

McDonald, F. J. & Allen, D. W. Training effects of feedback and modeling procedures on teaching performance. Technical Report No. 3, 1967. Stanford Center for Research and Development in Teaching, Stanford University.

Session	Group		
	DPF (10 S's)	PF (10 S's)	DP (8 S's)
Orientation	XXX	XXX	XXX
Pretest	XXX	XXX	XXX
Group Discrimination Training (Multiple Discriminations)	XXX		XXX
Practice (Multiple Discriminations)	XXX (6 children per trainee)	XXX (6 children per trainee)	XXX (3 children per trainee)
Feedback (Multiple Discriminations)	XXX (experimenter & videotape)	XXX (experimenter & videotape)	(peer & videotape)
Group Discrimination Training (Two-part Discriminations)	XXX		XXX
Practice (Two-part Discriminations)	XXX (6 children per trainee)	XXX (6 children per trainee)	XXX (3 children per trainee)
Feedback (Two-part Discriminations)	XXX (experimenter & videotape)	XXX (experimenter & videotape)	(peer & videotape)
Post-test	XXX	XXX	XXX

Summary of training and test sessions for the three experimental groups.

(SLIDE 1)

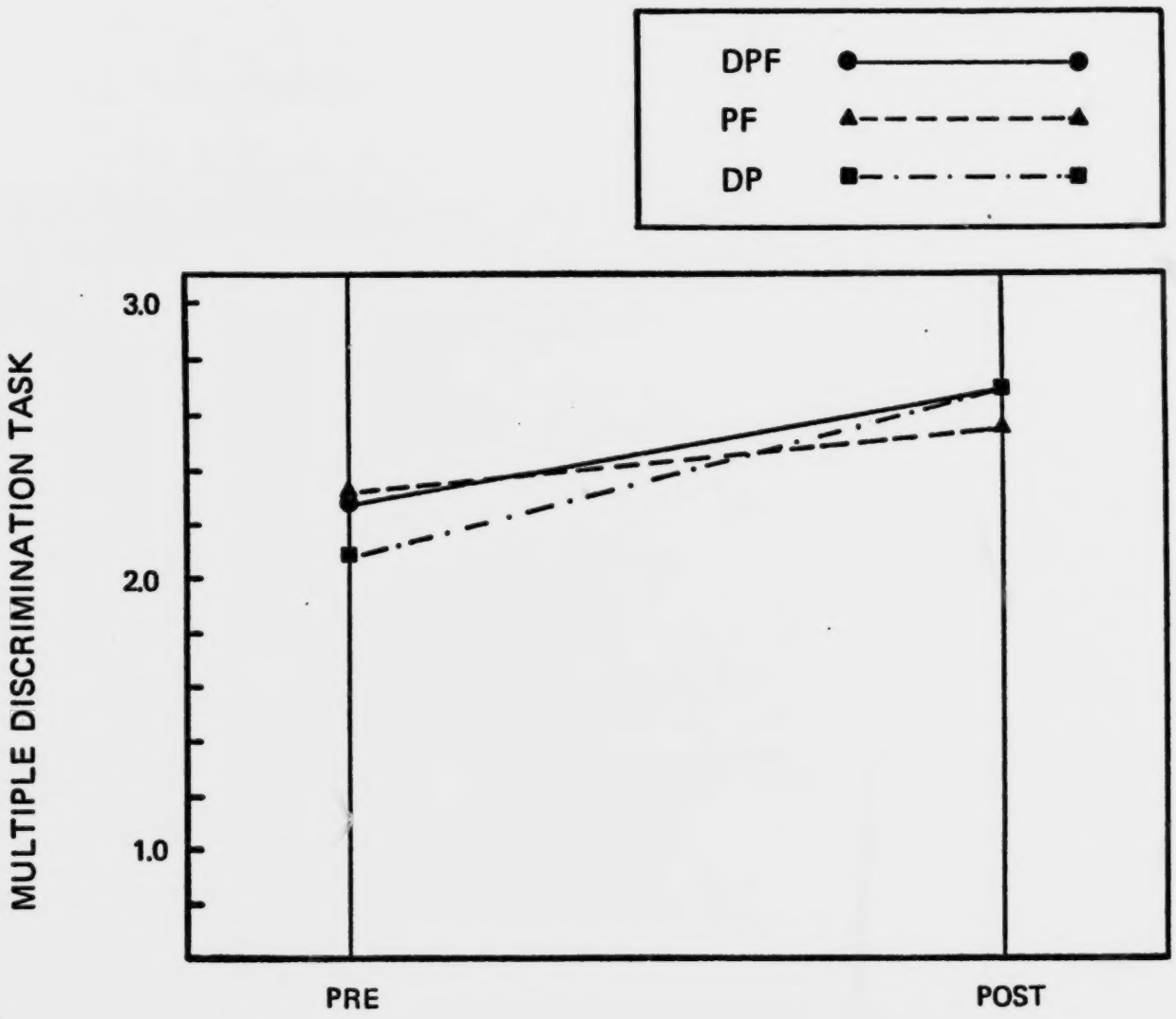
Objective Given: An unordered array of numerals.

Child can: Point to each numeral as it is named.

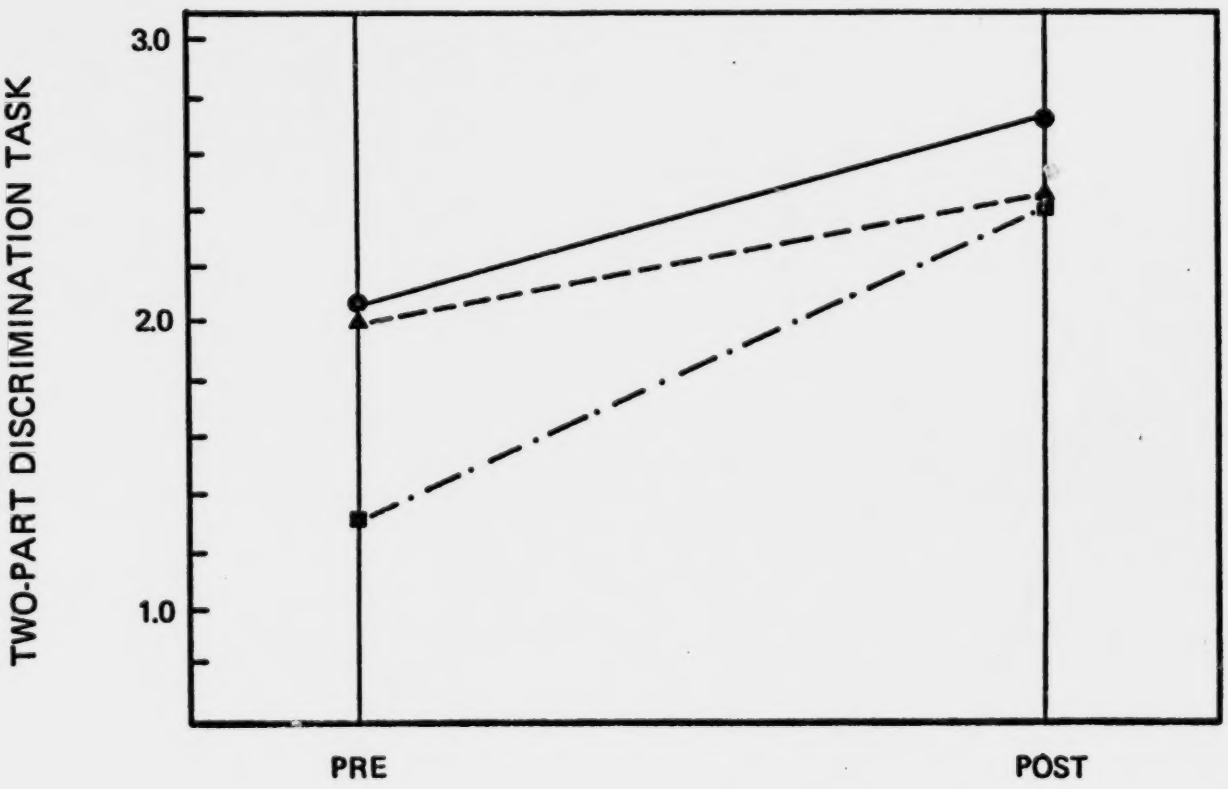
Numeral	Tally	Definitely cannot	Probably cannot	Possibly cannot	Can't tell	Possibly can	Probably can	Definitely can
1		3-	2-	1-	0	1+	2+	3+
2		3-	2-	1-	0	1+	2+	3+
3		3-	2-	1-	0	1+	2+	3+
4		3-	2-	1-	0	1+	2+	3+
5		3-	2-	1-	0	1+	2+	3+
6		3-	2-	1-	0	1+	2+	3+
7		3-	2-	1-	0	1+	2+	3+
8		3-	2-	1-	0	1+	2+	3+
9		3-	2-	1-	0	1+	2+	3+

Rating form for Certainty judgments

(SLIDE 2)

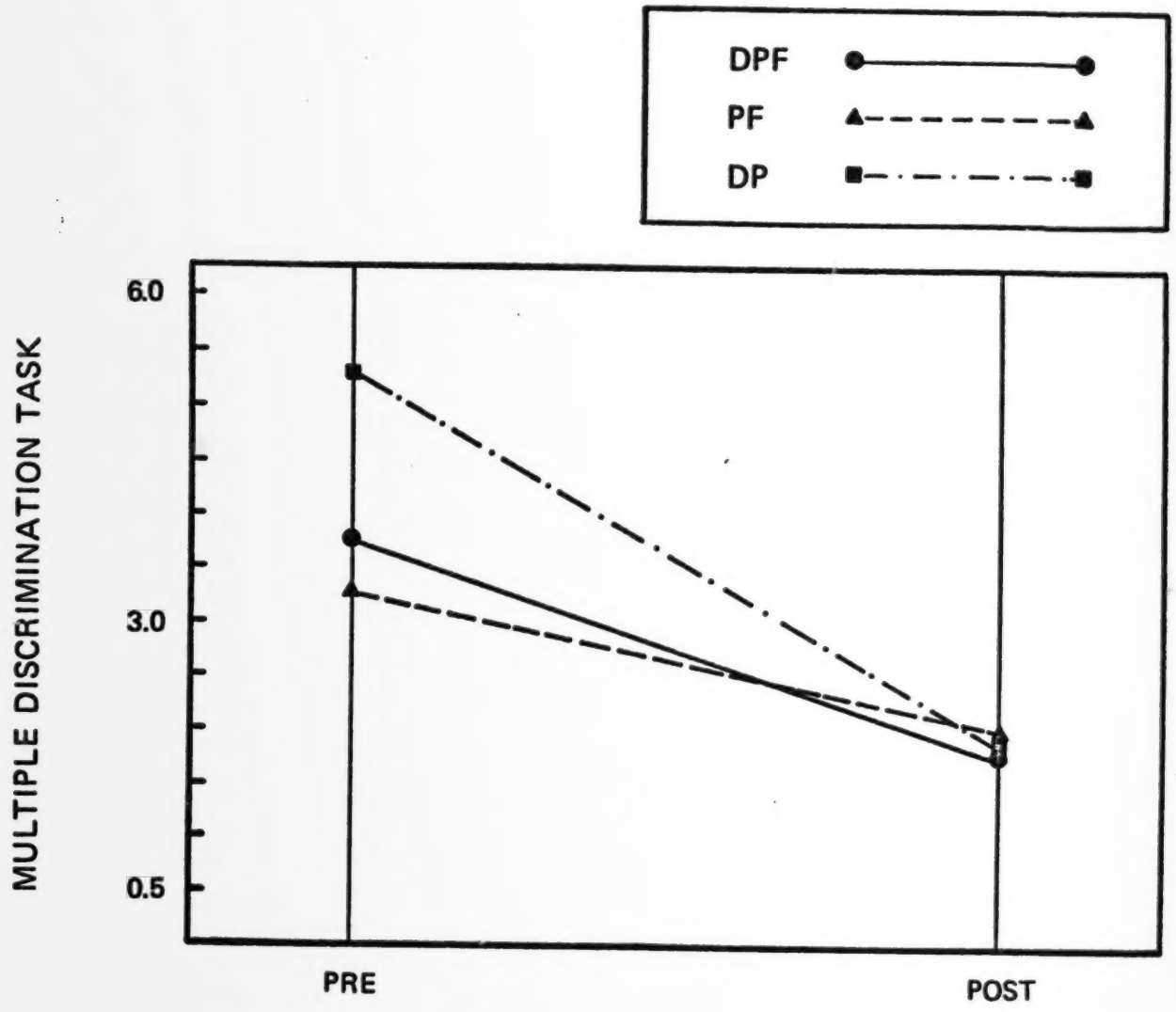


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

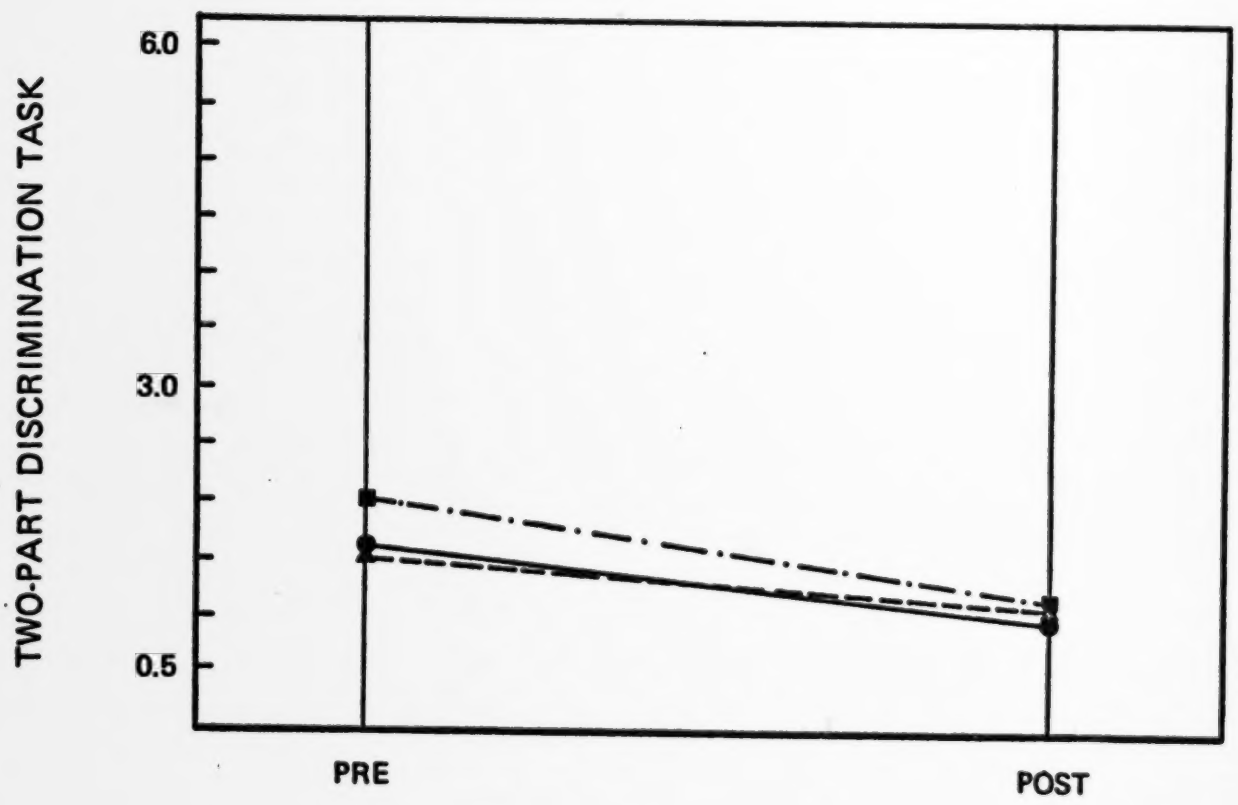


MEAN CERTAINTY RATING OF OBSERVERS ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 3)

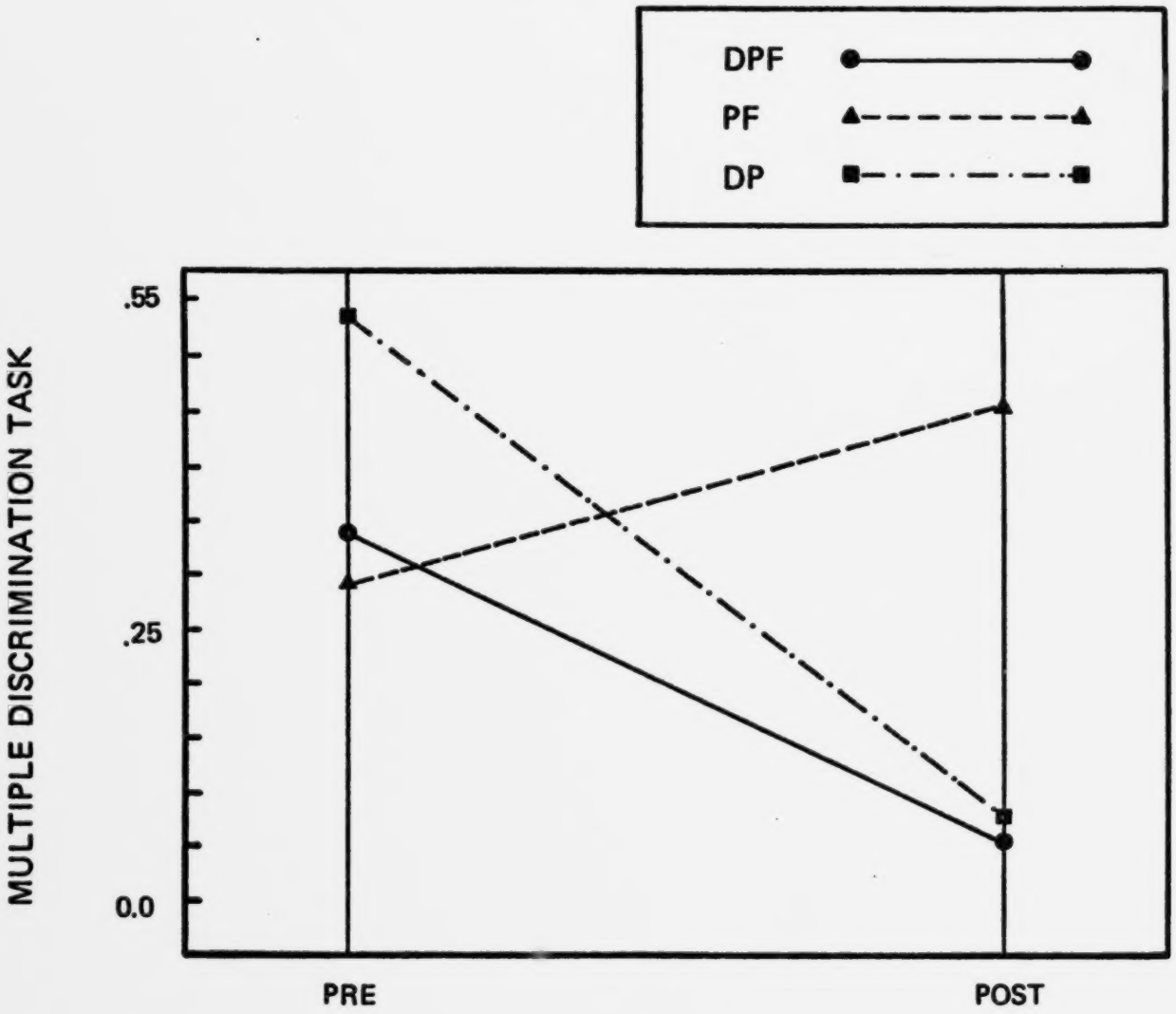


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

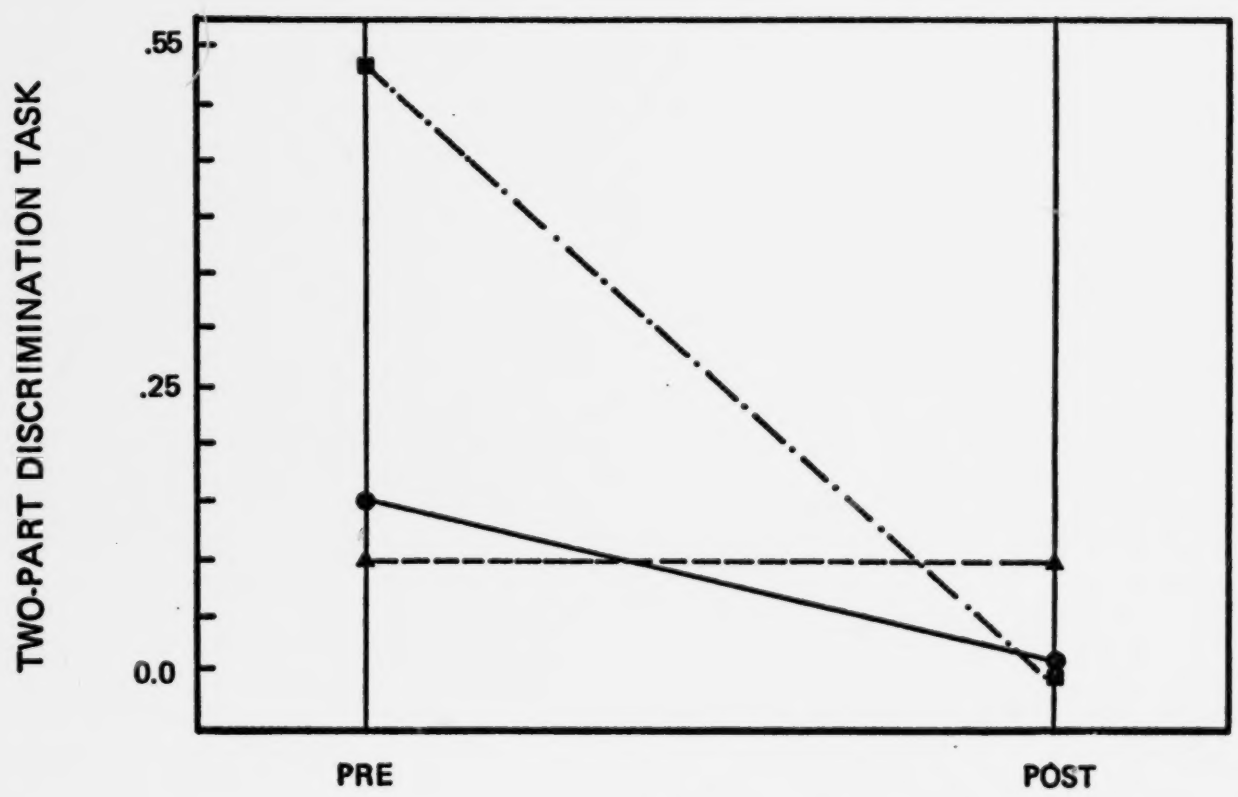


MEAN TIME ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 4)

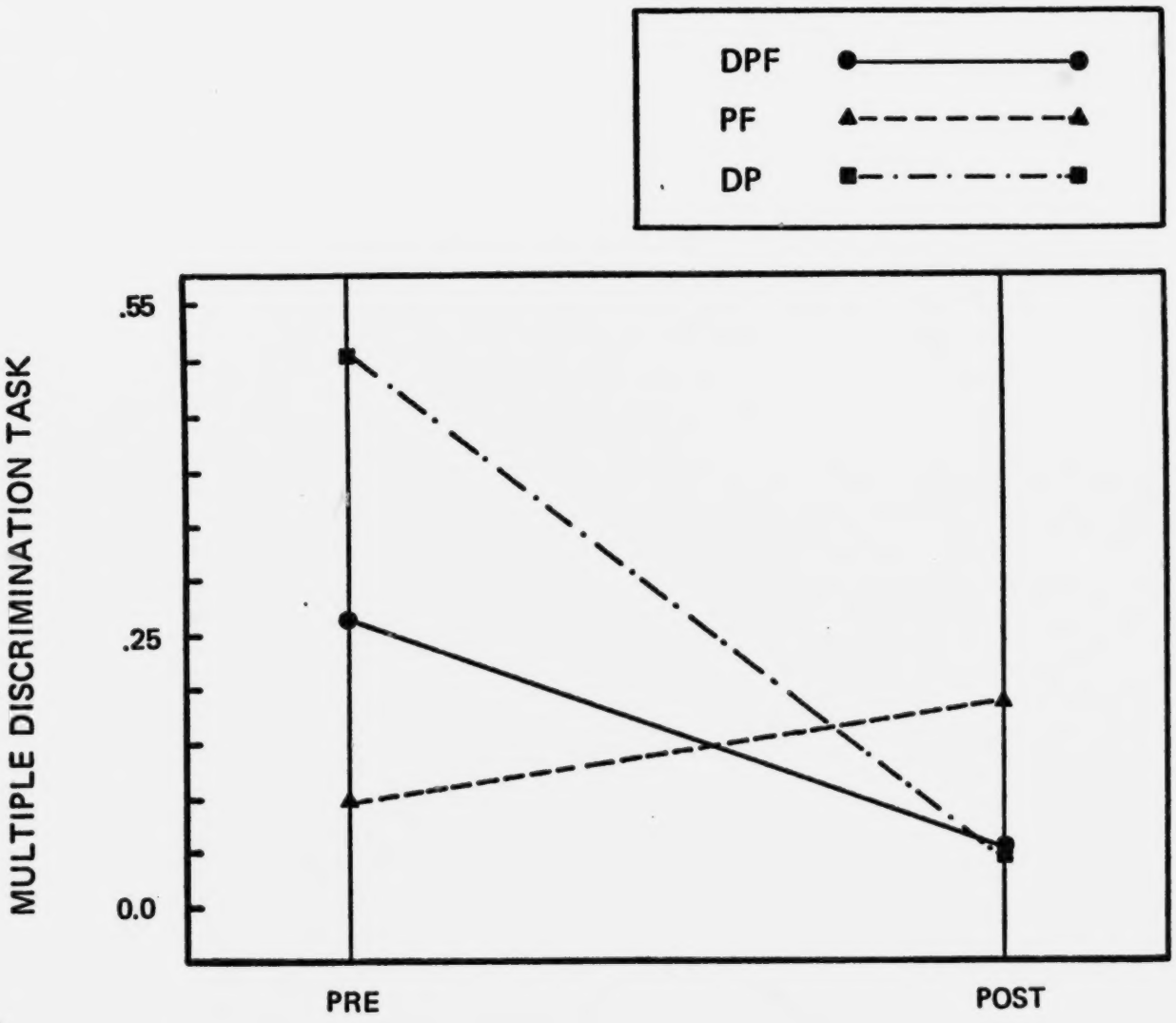


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

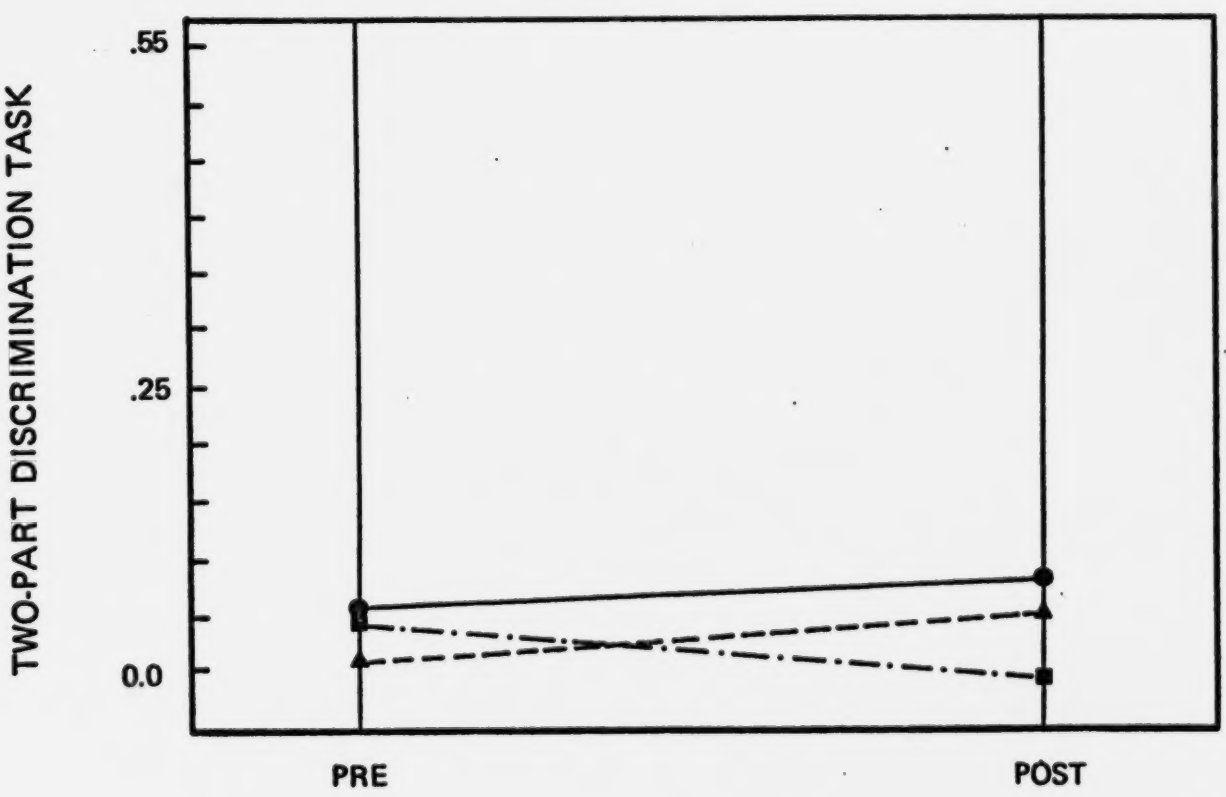


MEAN PERCENT OF TASKS IN WHICH TRAINEE ELIMINATES PART OF STIMULI ON PRETEST AND POSTTEST SESSIONS FOR EACH TREATMENT GROUP

(SLIDE 5)

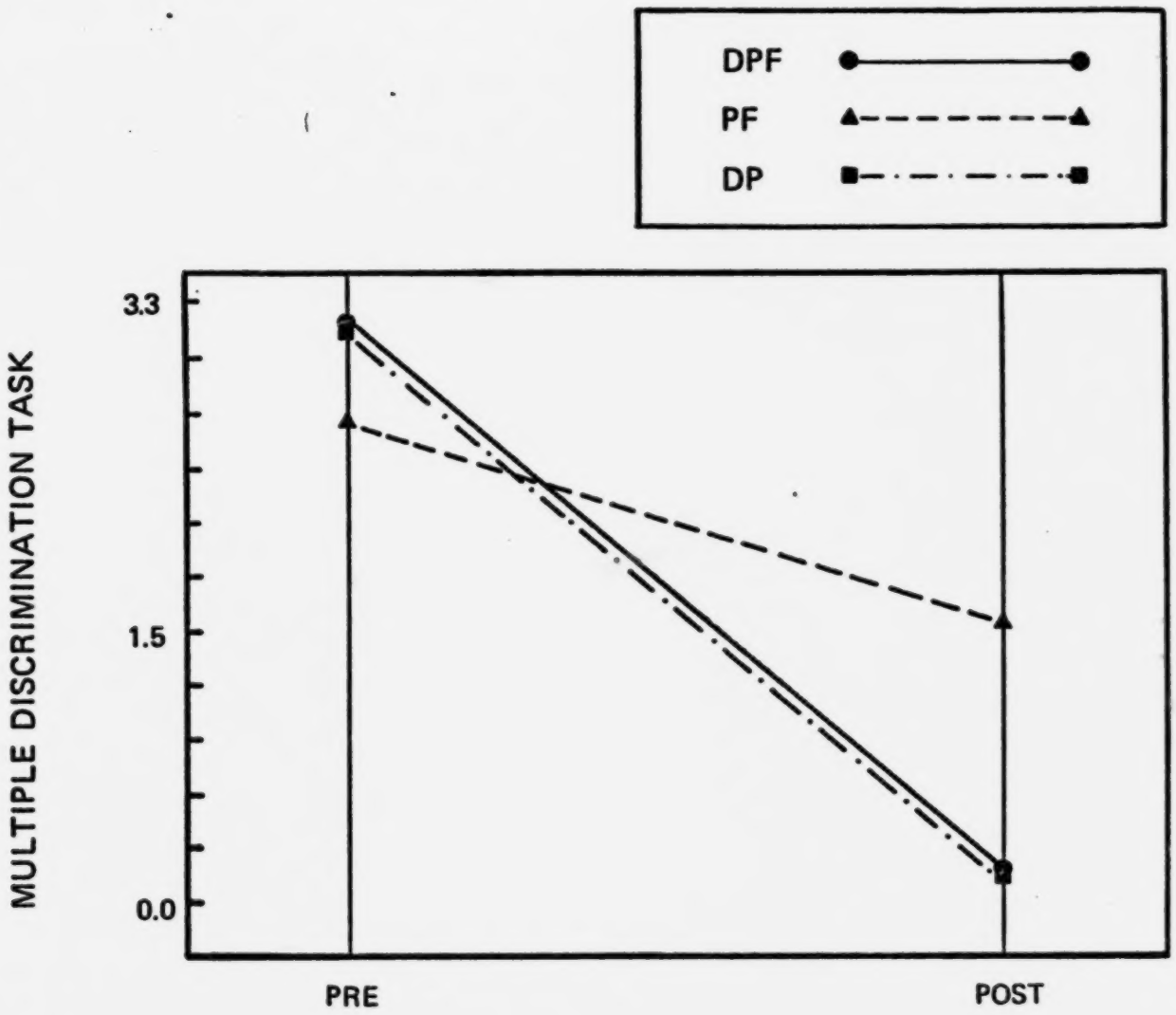


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

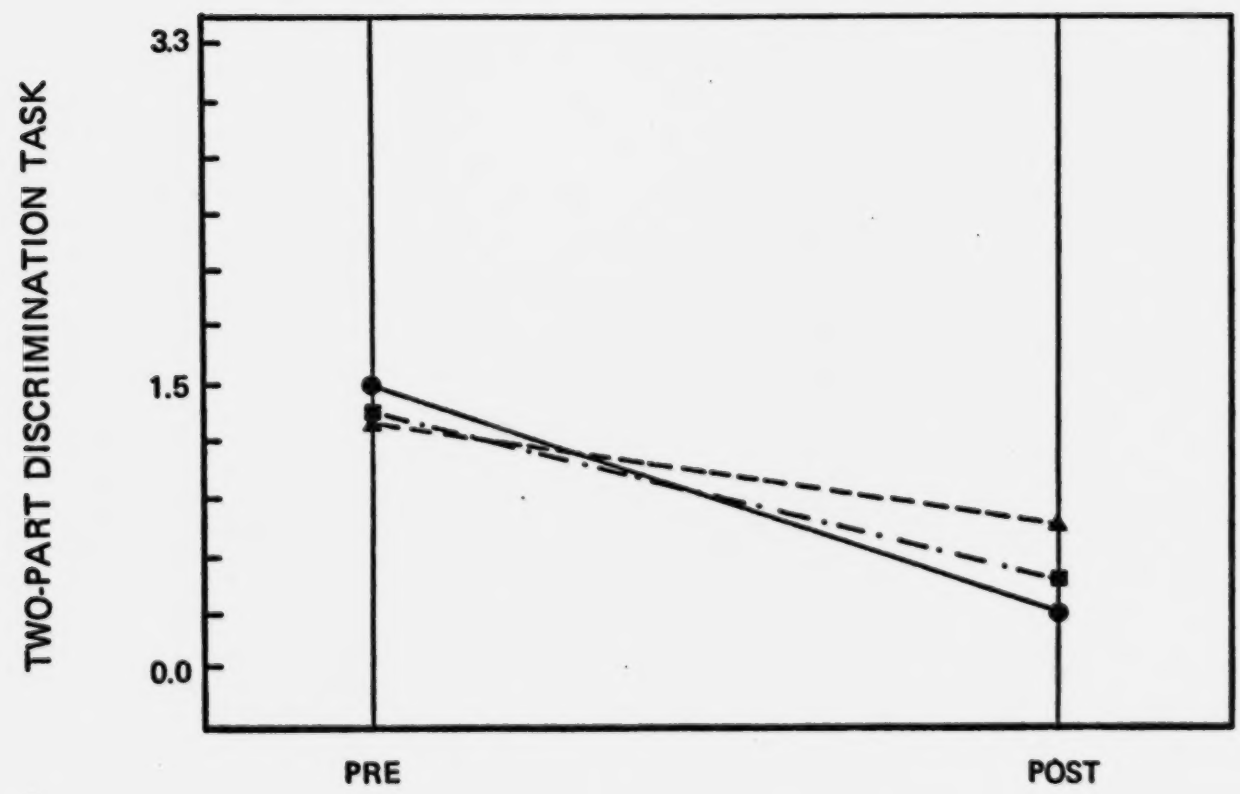


MEAN PERCENT OF TASKS IN WHICH TRAINEE ALLOWS CHILD TO ELIMINATE PART OF STIMULI ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 6)

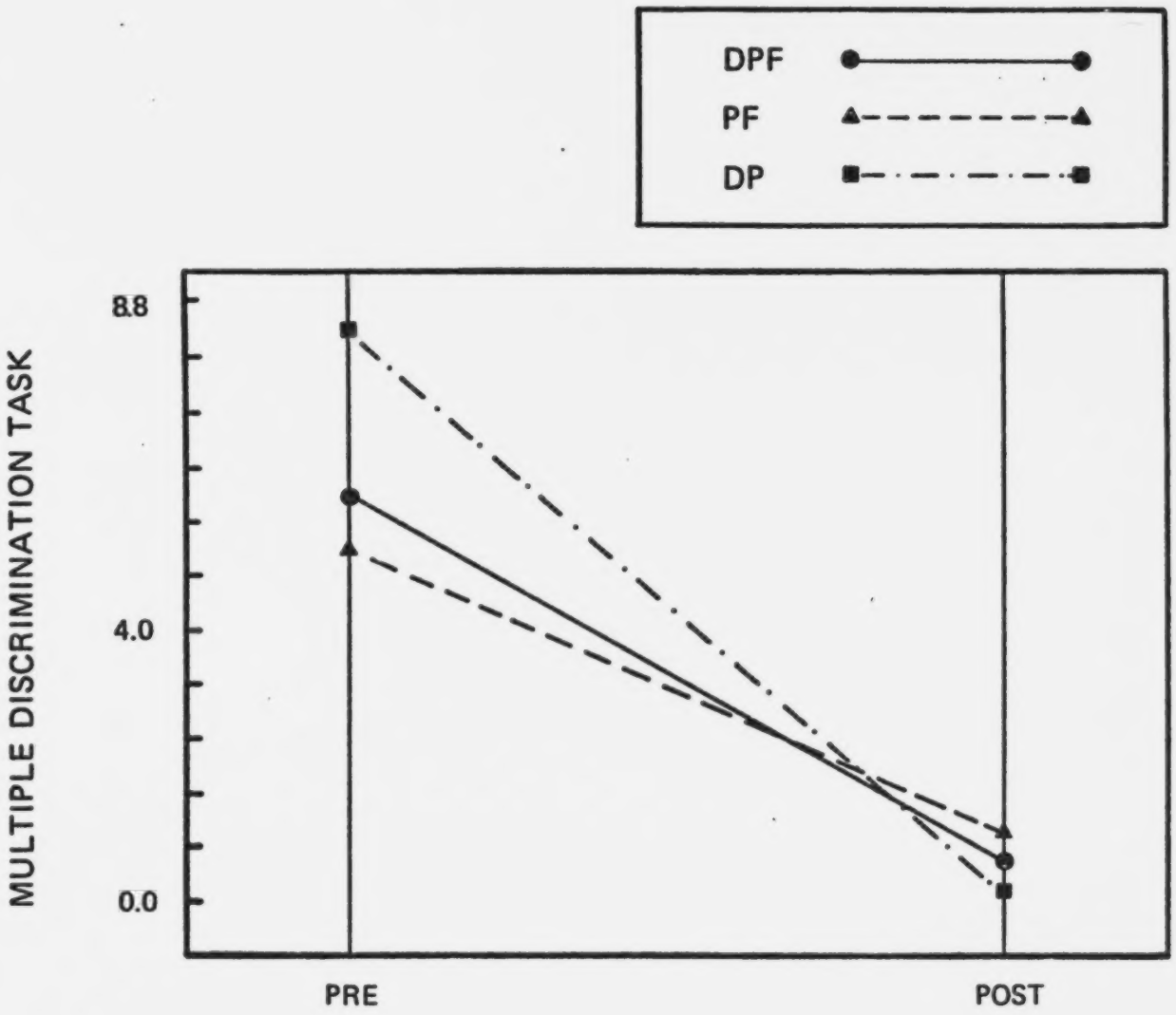


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

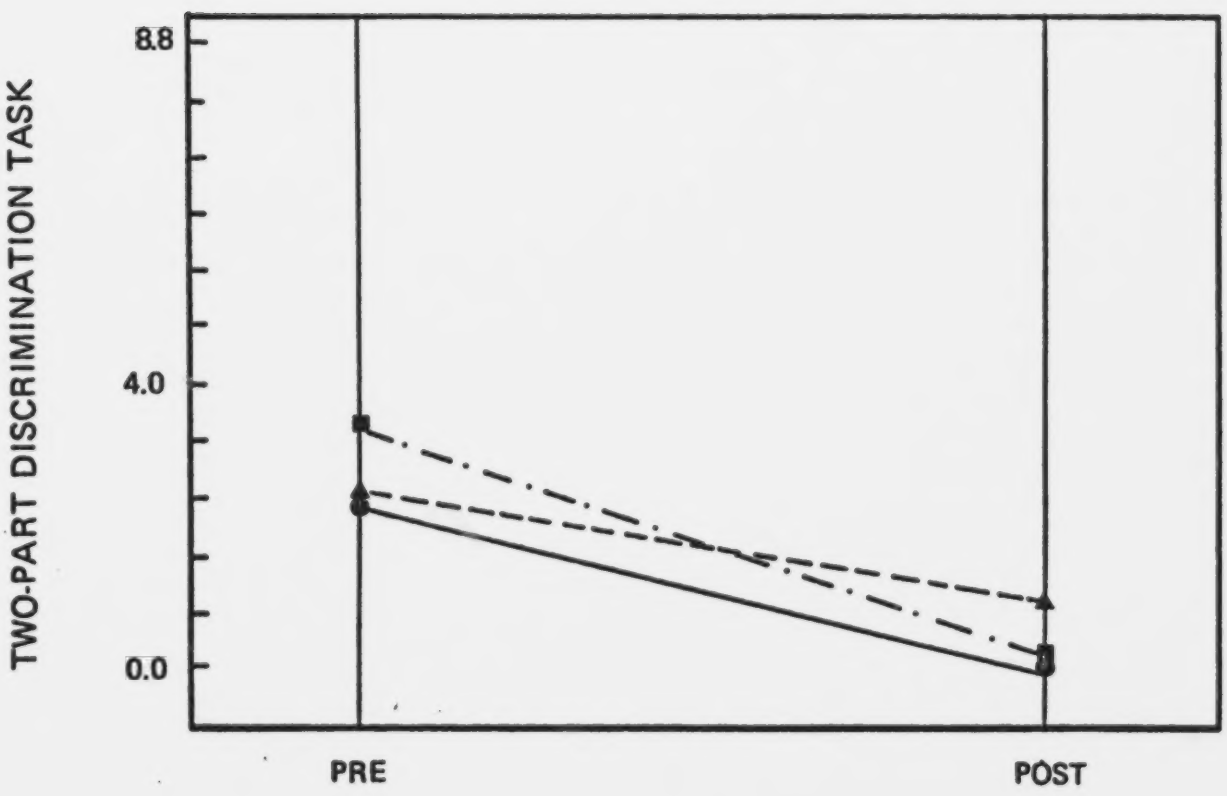


MEAN NUMBER OF CUES PER TASK ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 7)

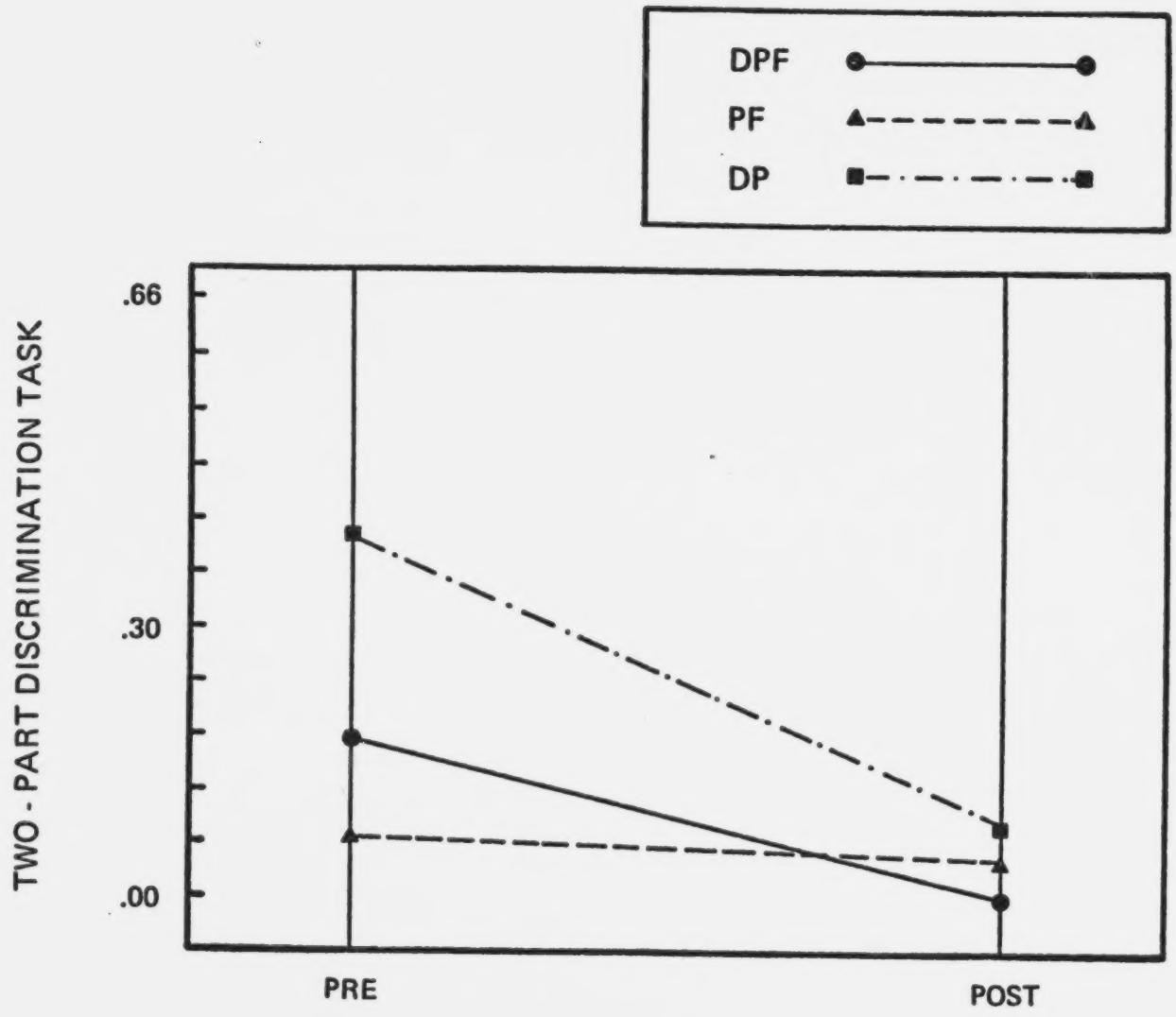


D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK

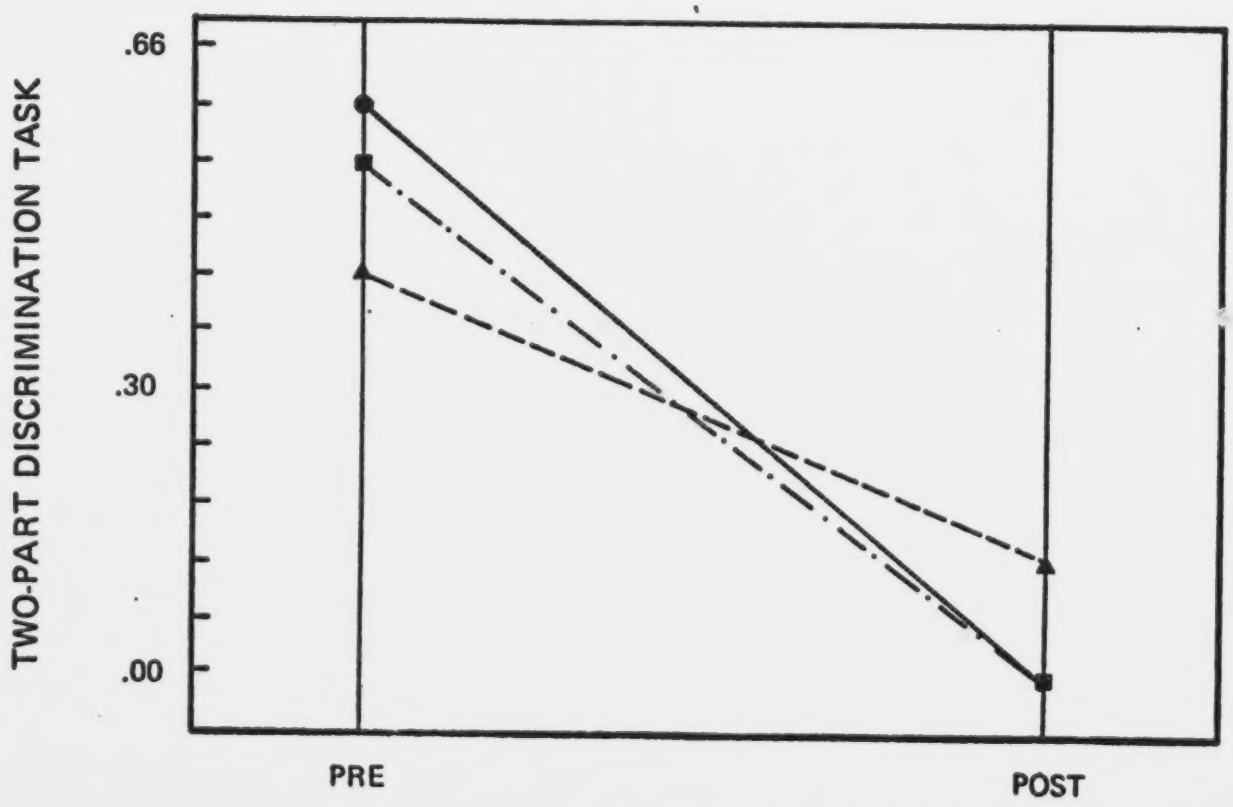


MEAN NUMBER OF VERBAL PROMPTS PER TASK ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 8)

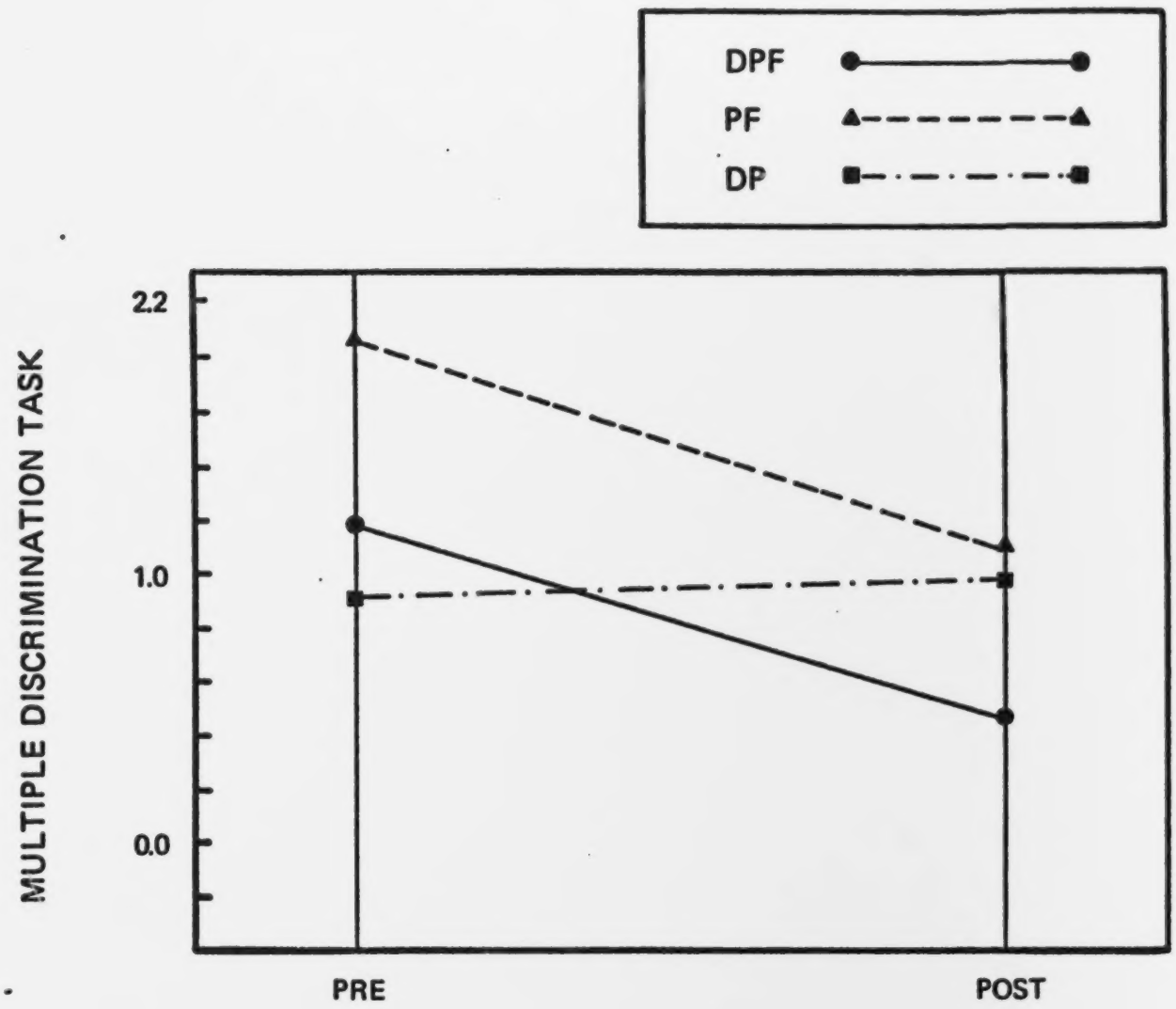


MEAN PERCENT OF TASKS IN WHICH TRAINEE FAILS TO ADD DISTRACTING MATERIALS BY PRETEST AND POSTTEST FOR EACH TREATMENT GROUP.

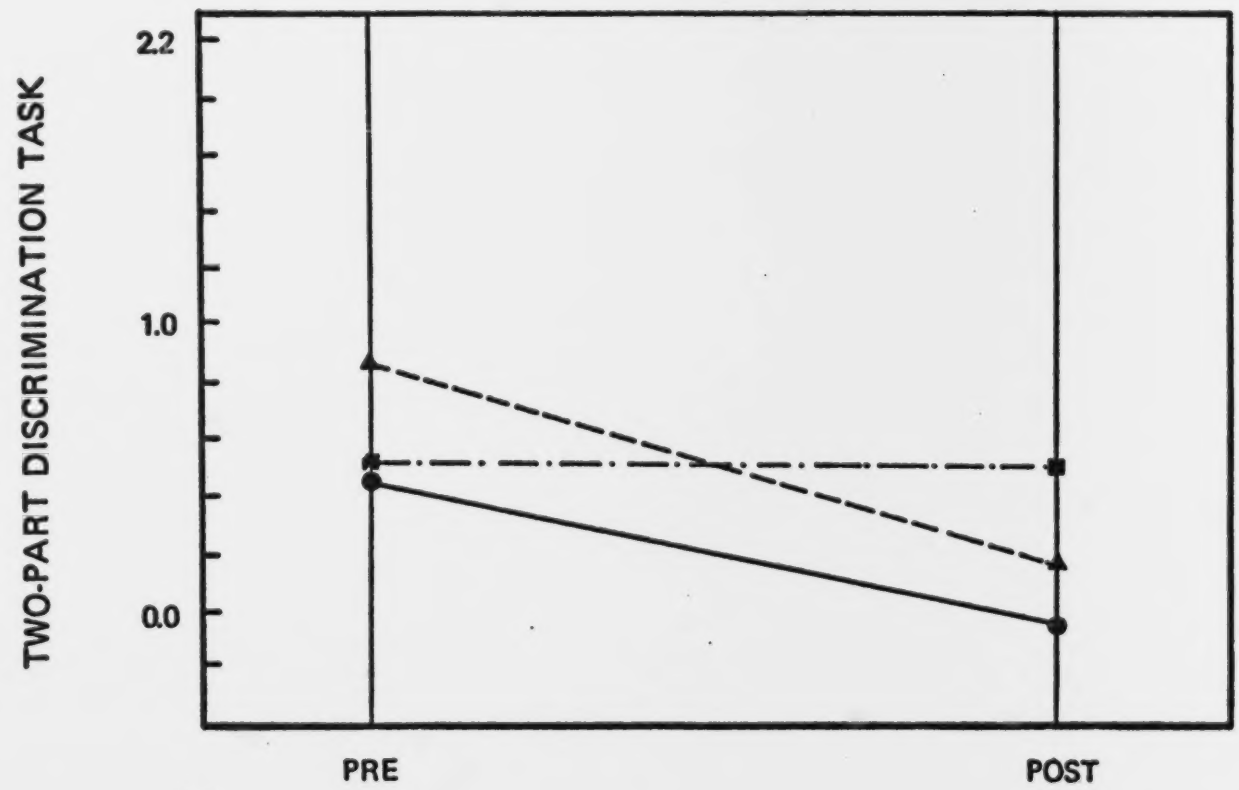


MEAN PERCENT OF TASKS IN WHICH TRAINEE FAILS TO DISTRACT CHILD BETWEEN RETESTS ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 9)



D INDICATES DISCRIMINATION TRAINING; P INDICATES PRACTICE; F INDICATES EXPERIMENTER FEEDBACK



MEAN ITEMS NOT RETESTED MARKED UNCERTAIN BY OBSERVERS ON PRETEST AND POSTTEST FOR EACH TREATMENT GROUP

(SLIDE 10)